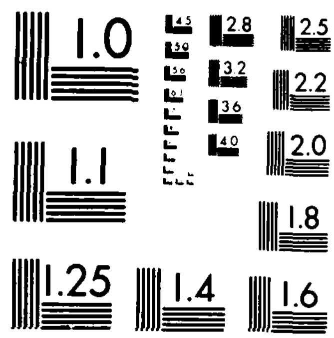
AD-A174 463 AN APPLICATION OF SIGNAL ANALYSIS AND PATTERN 1/1
RECOGNITION TO STUDY A SIMP (U) NEW MEXICO ENGINEERING
RESEARCH INST ALBUQUERQUE J M CARSON AUG 86 NMERI-20
UNCLASSIFIED AFOSR-TR-86-1076 AFOSR-82-0102 F/G 19/4 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD-A174 463

AFOSR-TR- 86-1076

(2)

NEW MEXICO ENGINEERING RESEARCH INSTITUTE

UNIVERSITY OF NEW MEXICO
BOX 25 UNIVERSITY STATION
ALBUQUERQUE, NEW MEXICO 87131

86 11 20 053

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION: UNCLASSIFIED

1b. RESTRICTIVE MARKINGS:

2a. SECURITY CLASSIFICATION AUTHORITY:

2b. DECLASSIFICATION/DOWNGRADING SCHEDULE:

3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; [illegible]

4. PERFORMING ORGANIZATION REPORT NUMBER(S): NMERI-20

5. MONITORING ORGANIZATION REPORT NUMBER(S): AFOSR-TR- 86-1076

6a. NAME OF PERFORMING ORGANIZATION: NEW MEXICO ENGINEERING RESEARCH INSTITUTE

6b. OFFICE SYMBOL (If applicable):

7a. NAME OF MONITORING ORGANIZATION: Air Force Office of Scientific Research

6c. ADDRESS (City, State and ZIP Code): CAMPUS P.O. BOX 25, UNIVERSITY OF NEW MEXICO, ALBUQUERQUE, NEW MEXICO 87131

7b. ADDRESS (City, State and ZIP Code): Bolling AFB, Washington, DC 20332

8a. NAME OF FUNDING/SPONSORING ORGANIZATION:

8b. OFFICE SYMBOL (If applicable): AFOSR/NA

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: Grant AFOSR-82-0102

8c. ADDRESS (City, State and ZIP Code): AN APPLICATION OF SIGNAL ANALYSIS (OVER)

10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
61102F	2307	C1	

11. TITLE (Include Security Classification): AN APPLICATION OF SIGNAL ANALYSIS (OVER)

12. PERSONAL AUTHOR(S): JAMES M. CARSON

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
Final Report	FROM Feb 82 TO Jul 85	August 1986	54

16. SUPPLEMENTARY NOTATION:

17. COSATI CODES

FIELD	GROUP	SUB. GR.

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number): GROUND MOTION, SIGNAL ANALYSIS, PATTERN RECOGNITION (OVER)

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

A simple problem involving the identification of an explosive source as being unbermed or bermed using a pattern recognition based analysis of buried ground accelerometer measurements is presented. This problem illustrates the advantages of computerized information extraction from the measured waveforms. Information was extracted from the frequency and cepstrum descriptions of the waveforms in addition to the more traditional time domain information. These signal features were incorporated into a Fisher's Linear Discriminant pattern recognition procedure. Previously unseen signals were classified with up to 100 percent accuracy depending on which features were used.

Close in explosive source measurements present unique problems to a pattern recognition based analysis approach. These problems are reviewed and approaches illustrated.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS ☐

21. ABSTRACT SECURITY CLASSIFICATION: UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE NUMBER (Include Area Code)	22c. OFFICE SYMBOL
Spencer Wu	202-767-4935	AFOSR/NA

DD FORM 1473, 83 APR — EDITION OF 1 JAN 73 IS OBSOLETE.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

11. (CONTINUED)

AND PATTERN RECOGNITION TO STUDY A SIMPLE GROUND MOTION PROBLEM

18. (CONTINUED)

FISHER'S LINEAR DISCRIMINANT
FREQUENCY DOMAIN
CEPSTRUM
SOURCE IDENTIFICATION

AN APPLICATION OF SIGNAL ANALYSIS AND PATTERN RECOGNITION TO STUDY A SIMPLE GROUND MOTION PROBLEM

James M. Carson

New Mexico Engineering Research Institute
Box 25, University of New Mexico
Albuquerque, New Mexico 87131

Final Technical Report
February 1982 - July 1985

Sponsored by
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
Air Force Systems Command
Bolling Air Force Base, DC 20332

August 1986

Research sponsored by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grant Number AFOSR-82-0102. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon.

CONTENTS

ILLUSTRATIONS

TABLES

TABLES (CONCLUDED)

I. INTRODUCTION

A program of high explosive testing has been carried out for some years to understand the soil and air transmission of the blast energy, the interaction phenomena of the blast event, and the response of structures that eventually must survive the loading; and to extend this understanding to the nuclear case. More complete analysis of these test results based on signal analysis and pattern recognition procedures can produce improved understanding of the controlling phenomena and should help influence the areas of test instrumentation, test design, and computer simulation.

While the application of signal analysis and pattern recognition techniques to this area of civil engineering does not guarantee improved results; signal analysis, often combined with pattern recognition; has been successfully applied and has led to major advances in areas such as cybernetics, bronics, non-destructive testing, radar, sonar, seismology, speech, brain wave analysis and photographic image interpretation and enhancement.

The principal objectives of this effort were

- To introduce pattern recognition techniques as an analysis tool to the civil engineering research area.
- To develop computer automated signal analysis procedures necessary to identify signal characteristics required in pattern recognition procedures.

In working toward these objectives the task can be conveniently separated into three broad areas: (1) data acquisition and signal processing, (2) feature extraction, and (3) application of pattern recognition procedures. Figure 1 illustrates the interrelationship of these areas.

Experimental data supplied by the Air Force Weapons Laboratory (AFWL) has been used to develop required acquisition, processing, and pattern recognition procedures. This report presents a number of solutions to a fairly simple two class ground motion problem involving the identification of the type of explosive source - a bermed versus an unbermed source. The various solutions presented are used both to meet the basic study objectives and to illustrate the unique characteristics of the problem and data encountered in this civil engineering research area. A second portion of this overall study, dealing with the indentification of ground areas that had experinced spall due to loading from an explosive event, is reported in a separate AFOSR Report (Ref. 1)

The sections that follow detail the characteristics of the physical problem, the data, and the way in which both adversely impact a statistically based pattern recognition procedure. Problem solutions illustrating the workings of the pattern recognition procedure are presented. Finally, the requirements for improved test planning, instrumentation and information extraction are discussed.

II. DATA BASE OVERVIEW

AFWL conducted a series of explosive tests dubbed PRE-HYBRID GUST (PHG) to better define pore air pressure effects. When soil is subjected to an airblast environment a phenomenon of interest is the expansion of the soil due to higher air pressures in the soil pores than at the surface. Ground motion is influenced by explosive energy transmitted directly through the ground and indirectly through the air (air slap).

The first three PHG events involved 13.6-kg surface tangent spheres of C4 explosive. The next two events were the same except that the C4 explosive was bermed to reduce the airslap. Table 1 outlines the PHG I series. Figure 2 shows the instrumentation arrangement. Of particular interest in Figure 2 are the 14 horizontal and vertical accelerometer gage pairs placed at various ranges from the surface explosive and depths below the ground surface.

Acquiring identical data sets is a fundamental problem. For instance, PHG events 1 through 5 exhibited crater size and volume differences and PHG events 4 and 5 had differing berm sizes. In addition the various accelerometers in a particular test were subjected to different physical phenomena. Other studies (Ref. 2 and 3) have shown that the gages relatively close to the explosive sensed ground spall and the associated rejoin. The following comments quoted frcm Reference 4 help illustrate differences observed in the manual analysis of the PHG data.

> GROUND MOTION
>
> . . . The first three events, PHG I-1 through 3, have ground motion data which are very similar in many respects. The first positive and negative peak accelerations are tabulated in Table [2]. The data from the lower gages, at depths of 0.76 m and greater, are very repetitive for these three shots. PHG I-2 and 3 at these deeper levels are more similar than PHG I-1 and 2 or PHG I-1 and 3. The upper gages at the 0.23-m depth did not show good reproducibility from shot to shot, especially at the ranges of 3.23, 5.67, and 6.55 m. The data from the gages at the 0.23-m depth for ranges for 8.50, 18.35, and 35.78 m are fairly similar for these three events.
>
> The two bermed events, PHG I-4 and 5, had similar data with good repetitiveness from shot to shot. The first positive and negative peak accelerations are also tabulated in Table [2]. One interesting difference

occurs at 35.78 m at a depth of 0.23 m: the horizontal acceleration for PHG I-4 is about 3 times the value for PHG I-5 (0.45 g and 0.13 g, respectively). Several of the traces from these events had high noise-to-signal ratios. This was caused by very low accelerations, especially vertical, as compared to the first three events. PHG I-4 and 5 were recorded using the same gages and calibrations as used for PHG I-1 through 3; therefore, large errors in quantizing and high noise-to-signal ratios were not unexpected.

Table 2 (Ref. 4) shows the first positive and negative acceleration peaks as well as judgments about data quality. Regardless of these quality judgments, all the data were used in this pattern recognition study.

PHG 1 through 5 were used as a data set to develop the data processing and pattern recognition procedures under investigation in this study. The pattern recognition procedure used in this study, and described later in this report, is effective in separating information into two classes. The two classes for this study are unbermed (PHG 1, 2 and 3) and bermed (PHG 4 and 5). On a physical basis one would look for the airslap component in the time waveform of the vertical accelerometer and use this to distinguish bermed from unbermed data records. It would be expected that the airslap-induced ground motion component in PHG 1-3 would be earlier and larger than that seen in PHG 4 and 5. It is reasonable to expect that this could be easily done for those gages close to the surface and close to the explosive. As gage range or depth increases this should become more difficult. It should also be remembered that this study is attempting to separate the unbermed and bermed classes within a data base that also exhibits crater size and volume differences, berm size differences, and spall/no spall differences.

III. DATA ACQUISITION

A rigorious and severe position was adopted in regard to data acquisition. Data coloration and interpretation were avoided to allow a more valid test of the pattern recognition procedure. The original PHG I-5 FM analog tapes were reread and digitized. The digitized portion of the signal consisted of approximately the first 1/2 second at a digitizing rate of 4096 Hz, producing 2048 digitized points. This is contrasted with the original digitizing rate of 50 kHz for 1/2 second (25,000 points). Digitizing at 4096 Hz means that the maximum frequency that can be detected is 2048 Hz, based on the Nyquist criteria. A check of the frequency content of the signals produced a maximum of less than 1000 Hz; thus a reasonable digitizing rate was used. Also, an antialiasing 1500-Hz, 5-pole constant delay (CD) or Bessel low-pass filter was applied to the data..

It should be noted that the original reports describing the PHG series refer to corrections such as filtering, removing noise spikes, and appropriate baseline corrections. Also, it was noted that some of the reported data were inverted because they did not "look right." No doubt some of these corrections are valid; however, as is often the case, the specific criteria for and the details of the corrections and inversions were not reported. Other than the antialiasing filtering already mentioned, no data correction of any kind was performed in the acquisition phase of this project.

Following the acquisition of each digitized signal, the signal and the frequency transform of the signal were displayed and a check of the digitizer quantum range made to ensure that an adequate range was used and that the maximum range was not exceeded, which would have clipped the signal. Figure 3 shows a typical signal acquired. The MIN and MAX refer to the digitizer range of this signal. The analog-to-digital converter range is ±2048. In addition, signal calibration records for each transducer were recorded. The signal associated with the fire pulse of each explosive event was also recorded. Examples of these signals are shown in Figures 4 and 5. Although rereading the numerous FM analog records was time consuming, it was believed to be a necessary task because it was important that the processing history of the signals be known.

The data acquisition effort was completed on the AFWL Signal Analysis System (Ref. 5). This system consists of a PDP 11 computer coupled with a General Radio analog-to-digital converter and various pieces of analog front-end equipment to allow the reading of FM analog tapes. Because the language of this system is

Time Series Language (TSL), the system was not capable of some of the mathematics required in the pattern recognition phase of the effort. The digitized signal data were transcribed to a FORTRAN-compatible format (ASCII) and transferred to the Civil Engineering Research Facility (CERF) PDP 11 computer, where the various signal processing and pattern recognition steps took place.

Although Table 2 notes various poor quality, bad, or even missing records, it should be noted that all data regardless of quality was used in this study. Also, no data records were found to be missing. Following the acquisition of the data a visual examination of the history waveforms did show some "poor" quality records. Poor quality here means a low signal to noise (s/n) ratio or a shape characteristic that from experience may indicate a faulty gage response. Table 3 lists these poor waveforms by test and gage member and notes if the records were normally used in the pattern recognition learning or test sets. Examination of the prediction results were made to see if these records were consistently misclassified. Figure 6 shows examples of these poor records, comparison with the good quality record previously presented in Figure 3 shows just how poor these records are.

IV. SIGNAL PROCESSING

The digitized data underwent several processing steps prior to the application of the Fast Fourier Transform (FFT) and feature extraction procedure. These steps are listed below.

1. The digitizing of each signal was started a short time (10 ms) prior to the explosive fire pulse (fidu) that serves as a zero time baseline. The digitizing of the fire pulse was started at the same time as the accelerometer signals. The number of points from the start of the recording to the fire pulse was counted and subtacted from the start each signal record. This produced a consistent time baseline for all records. This process was computerized and thus automatic.

2. The early time portion of each signal record (30 points prior to the fidu) was averaged. Because this was a portion of the record where no signal was present, and thus should have a zero mean, the computed average was used to adjust the d.c. level of the entire signal. It is believed that this type of d.c. correction is unambiguous and clearly justified. No other types of d.c. corrections, such as ramp corrections, were attempted.

3. The transducer calibration was applied to each signal to transform the signal amplitude from arbitrary quantum units (integers) to physically meaningful engineering units (reals). This step is not required in all cases because data are often normalized. Also, there is sometimes utility in dealing with integer quantum numbers.

Following these steps, the data were frequency-transformed via an FFT, and features were extracted from both the amplitude-time, frequency, and cepstrum descriptions of the signal. A feature is any characteristic of the signal or transformed signal that is believed to have some significance. A computerized feature extraction procedure was developed and was used to extract the features listed in Table 4. Note that 30 features were extracted from both the vertical and horizontal accelerometers. It should be noted that the features listed in Table 4 are quite simple and that those in the time domain are similar to those signal characterics often applied in the usual manual interpretation of data.

The transformations used were the Fourier Transform (via an FFT), which presents a signal amplitude vs. frequency representation of the information in the signal and the power Cepstrum (Ref. 6) which is the fourier transform (power spectrum) of the logrithm of the fourier transform. The Cepstrum highlights the

time difference between superimposed portions of signals. This ability to sense "echoes" was developed to aid the solution of the spall detection problem reported in Reference 1, but was also applied to the berm problem.

The selection of features in general can be arbiturary, although physical knowledge and insight are useful. Basically, however, the signal must contain information of the phenomena of interest if signal analysis is to help extract it. No new information is created. Transforms and signal analysis procedures only make particular aspects of the information more apparent. This implies that there is no substitute for intelligent and imaginative experimental procedures to acquire quality data.

V. PATTERN RECOGNITION

The unbermed/bermed two-class problem was solved using a Fisher Linear Discriminant (FLD) pattern recognition procedure (Ref. 7) described in the next section. This procedure uses the d-dimensional feature vector describing each signal produced in the feature extraction process. A representative set of signals from each class, called the learning set, is selected. Weighting factors are determined for each component of the vector. These weighting factors project the vector to a line in space so that the separation of the two classes lying on this line is maximized. This unique line is defined by a mathematical process that maximizes the distance between the two class means while minimizing the within-class scatter (variance) of each class. This process involves finding the inverse of a matrix. This matrix is usually nonsingular when the number of data samples in the learning set exceeds the dimension of the feature vector. Limitations in the size of the learning set may force limits on the dimension of the feature vector used at any one time.

The application of Fisher's Discriminant involves selecting a subset of data to be used for determining the weighting factors. This set is called the learning set, and the correct classification for these data is fed to the computer. To evaluate the performance of the FLD procedure, a previously unseen set of data, called the test set, must be used. The weighting factors are applied to these data, and class predictions are made. Comparing the predicted to the actual class allows an evaluation of the FLD performance.

In the basic unbermed/bermed study one-half the data from each of the five events were randomly selected and assigned to the learning set. The remaining data were assigned to the test set.

For this study the same features were extracted from both the horizontal and the vertical accelerometer gage records. These features were then stacked into a single feature vector. Thus each feature vector can contain information from a gage pair.

The structure of this problem is best seen by considering a matrix of n columns and d rows (Fig. 7). In this particular problem a particular gage pair (horizontal and vertical accelerometer) at a specified range and depth was considered a data sample. This was necessary in order to have a reasonbly large number of columns. In the normal studies this produced 35 (1/2 x 5 tests x 14 instrument pairs/test) columns consisting of both bermed (n_2 = 14) and unbermed

(n_1 = 21) data in both the learning and test sets. This strategy introdued an additional "test" difference --that of range from the explosive-- into the data base. Thus the signal features chosen and the pattern recognition procedure had to be capable of detecting the difference between a bermed and unbermed source while ignoring the amplitude and arrival time difference associated with the range differences inherent in the chosen data structure.

The rows of the matrix are composed of the 30 features mentioned (Table 4) for both the horizontal and vertically sensitive gages resulting in 60 rows.

The number of rows easily exceed the number of columns. Thus in order to avoid a singular matrix problem only portions of the available features can be used at any one time. It is thus clearly desirable to find techniques to identify the "good" features and eliminate the "bad" features. A number of studies, reported in a later section, explored techniques to identify good features.

It might be noted that in many pattern recognition problems a few features are ultimately identified that yield consistent classification results. Presumably these features are good physical descriptions of the phenomena. No rigorous procedures exist to separate the superior features.

VI. FISHER'S LINEAR DISCRIMINANT

Fisher's Linear Discriminant (Ref. 7 and 8) produces a 2 class decision line in d-dimensional space for d features considered simultaneously. This unique line is defined by a mathematical process which produces the optimum separation of the two groups or classes of projected data by maximizing the separation of the projected means relative to the scatter (variance) of the individual groups. This procedure reduces the dimensionality of the feature space and produces a more manageable problem. The mathematics of this procedure involves finding the inverse of a matrix whose rows are features and whose columns are instruments or tests. A problem exists here for the case of small data sets (small number of instruments/tests). If the number of columns exceeds the number of rows a non-singular matrix will usually result.

A contrived example of the value of this procedure in two dimensional space is shown in Figure 8. The problem is to identify "X" and "O". Neither feature 1 nor feature 2 could independently successfully classify "X" and "O". The Fisher Linear Discriminant, however, would calculate the line in space shown that would produce the best separation between the "X"s and "O"s (shown as solid circles and squares when projected to the line).

This procedure is normally applied to two class (group) problems and requires the data be broken into two sets - a learning set and a test set. The learning set and the associated correct groupings are used to establish the line in space and calculate the weight vectors with project each d dimensional data point to the line. The test set of data is then classified using the same weight vectors and the success of the prediction determined.

Since the FLD is statistically based (mean and variance) it can be subject to all the problems of small data sets, including the effects of outliers. Partly as a result of this, the following three modifications were made to the standard FLD:

MOD 1. Correction of Variance Pooling.

The within-class scatter matrix, $Sw = S_1 + S_2$ is reported (Ref.7) proportional to the covariance matrix for the pooled d-dimensional [feature] data. This is only true if the number of class 1 data is equal to the number of class 2 data ($n_1 = n_2$), however. Proper variance pooling was inserted in the FLD

$$Sw = \frac{n_1 S_1 + n_2 S_2}{n_1 + n_2}$$

as modification 1.

MOD 2. Modification of Data Mean.

The median value of a data set is nearly insensitive to outliers. Thus, the difference of the class mean feature vectors $(\bar{m}_1 - \bar{m}_2)$ was replaced with the difference of the median feature vectors $(m_1 - m_2)$ as modification 2.

MOD 3. Elimination of Outliers.

Modification 2 addresses the effect of an outlier on the mean but does not solve the problem with respect to the variance or scatter. The problem of outliers can fall into two categories. First an outlier can be a stray feature value, perhaps due to the feature extraction algorithm being insensitive to a particular waveform. Secondly an outlier can be a bad data point. A data point could range from the output of a single instrument to the output of an entire experiment. The following procedure involves routinely discarding the K highest and lowest data points from each class for the calculational portion of the FLD. These discarded points are flagged and retained for predictions, however. The drawback of this procedure is that it further reduces already small data sets.

Prior to the FLD weighting factor calculation a data screening procedure is performed within each class.

a. calculate feature mean

b. calculate feature variance

c. find normalized class matrix values

$$Vij = \frac{(Xij - \bar{X}i)^2}{Sii^2} \qquad \begin{array}{l} i \text{ - features} \\ j \text{ - data columns} \end{array}$$

d. sum the normalized feature vectors $\sum_{j=1}^{d} Vij$

e. flag and discard the largest K values

f. proceed as before with the reduced data set.

This procedure is sensitive to both outlier data points and features. As the number of features in the feature vector increases, it is suspected that this procedure becomes more sensitive to outlier data than to outlier features.

VII. FEATURE SELECTION

A number of procedures were used to help select those features or feature combinations containing "useful" information or discard those features that are redundant or without problem discrimination value. It has already been noted that no rigorous procedures in this area exist. This section will present background and show examples of the various procedures attempted. Later sections reporting test results discuss the value of these various feature selection procedures.

Outliers

An outlier is a data member of a set which appears not to belong to that set. Procedures to identify and discard outliers are usually ad hoc. It appears that "outliers" are too frequently discarded without justification. If statistical methods are used to identify and discard outliers they (the discard criteria) should include the number of points in the set and some measure of the set distribution (i.e. standard deviation). An implicit assumption, hard to prove for small data sets, of a normal distribution often occurs.

The effect of an outlier, particularly for small data sets, is to shift the mean and increase the set variance. In a statistically based pattern recognition procedure these effects can be important.

Modifications to the Fisher linear discriminant, already described, were made to adjust for the presence of outliers. It appears that two basic approaches are possible for outliers. The first approach involves deleting either features (rows) or instruments (columns) that appear anomolous. The difficulty of eliminating features is that a different feature would tend to be eliminated form each column thus leaving a matrix full of holes. If the variance of each feature in a feature vector is normalized and summed then the variant columns (tests) can be identified. This process was used in the FLD modification.

If it were desired to delete anomolous feature values a second approach involving filling in the deleted or missing data would have to be taken. To achieve this approach the data base should be a statistically based experimental design to facilitate the required interpolation. Available data is not usually in this format.

Probability Density Function

A probability density function - a graph showing the probability of a specific feature value - can be estimated from a data set. For small data sets the uncertainity associated with the PDF increases. If the PDF for the two classes of interest and a single feature are superimposed (Fig. 9) a visual comparison can help select features capable of separating the problem classes. Figure 9 shows some separation is possible. This is the usual situation and illustrates the need to use multiple features to produce the required class separation.

Scatter Plots

A scatter plot is simply the values of one feature plotted against the corresponding gage values of a second feature. If two classes are involved each class can be plotted as a different symbol as shown in Figure 10. This technique allows a simple visual analysis of class clustering due to the two features, of the existance of outlier values and of the distribution of the feature values. Figure 10 shows a possible outlier and Figure 11 shows a good partial separation based on feature 140.

Linear Correlation Coefficient

Correlation coefficients address the similarity of information in various features of a particular data set. Correlation coefficients were calculated for every combination of 2 of the 60 possible features for all the data and for each class (bermed and unbermed) of data. High correlation coefficients (near 1) between two features indicate similar information in the two featuress and thus both may not be needed. Low correlation coefficients (near zero) show that the two features are not related but do not directly indicate the merit of these features. In examining the results of the correlation coefficient study in light of the scatter plots it was found that a single "outlier" in the data could change a correlation coefficient from a very low number (.000013) to a high one (.910863), again illustrating sensitivity to "outliers".

If two features contain the same information one may be eliminated from the feature vector or the two could be combined to give a more reliable feature. This procedure may also give some physical insights about the interrelationships between the various features chosen.

Finally a matrix was made showing each feature against all others. The high ($r \geq .7$) and low ($r \leq .3$) correlation coefficients were recorded as "X" and "O". The resulting patterns showed groupings of similar and dissimilar feature information.

VIII. PROBLEM LIMITATIONS

While the data base used in this study was selected to be amenable to a pattern recognition attack it suffers from problems common to much of the nuclear simulation data.

1. Sparce Data Base - Nuclear simulation data often consists of only one shot and may be instrumented such that there is little relationship or redundancy of the instruments. Multiple shot series where the test parameters and instrumentation layout are consistent are rare. The resulting small data bases affect the size of the feature vectors one can investigate. Also, since many pattern recognition procedures are statistically based small data sets are a disadvantage. The PHG test series is actually relatively robust. Nevertheless the necessary ploy of considering each accelerometer pair as a test introduced additional problem complexity.

2. Nonstationary Data - Many signal analysis procedures assume a stationary signal (zero mean). The impulsive nature of explosive events tends to produce data records characterized by a sharp initial spike followed by an exponential decay - a signal that is clearly not stationary.

3. Nonergodic Data - Signal processing procedures often assume ergoditicity, that is that any time increment of a signal has the same frequency content as any other time increment of the signal. The decaying exponential of explosively generated signals does not fall into this category. Furthermore, signals propagating through the ground are dispersive and thus clearly not ergodic.

4. Noise - Noise is present in most experimental data and particularly present in explosively induced ground motion data. Noise generally makes it difficult to apply sophisticated signal analysis procedures such as the complex cepstrum or deconvolution procedures. Noise can often be reduced by signal averaging, however, this is not a reasonable option in this case due to the small data set and due to the lack of redundancy in the experimental design.

IX. FISHER LINEAR DISCRIMINANT PROBLEM SOLUTIONS

A large number of Fisher problems were investigated with the purpose of first demonstrating successful solution of the berm/no berm source classification problem and secondly to understand and illustrate important characteristics of the solution and to learn more about the nature of the problem, features used, and the information in the measured waveforms. Selected problem solutions will be presented in the following tables to illustrate various characteristics of the problem and solution techniques. In general the various problem solutions consisted of using different numbers and combinations of features, using features selected on a particular basis such as the correlation coefficient, using different data in the learning and test sets, and using the different versions of the FLD. Test numbers in the tables are simply a convenience for reference purposes.

Figure 12 shows the PHG gage arrangement for the five experimental tests. The first three events (A,B,C in Figure 12) were not bermed and the last two (D, E in Figure 12) were bermed. The circled letters show the data that was used in the standard learning set. One half of the data in each event was randomly selected to belong to the learning set. The remaining half of the data was placed in the test set. Note from Figure 12 that this procedure resulted in all gage pair locations, except location 315, having some representation in the learning set. However, each location does not always have learning set representation for each of the two classes. This learning/test set selection is the baseline used throughout the FLD based studies. Variations were made from this baseline to investigate the nature of the problem and will be identified in the following discussion.

Basic Studies

A number of FLD tests were conducted to show the basic feasibility of using various features and a pattern recognition procedure to classify the no berm/berm experimental data. Table 5 summarizes these results. The format of Table 5 will be used in discussing the other study results and is described in the following paragraph.

The study tables list the test number, summarize the features used in the study (FEATURES), the study learning and test set description and the prediction results for class 1, class 2, and the classes combined. The learning/test

criteria (LTC) code needs further explanation. "L" stands for the standard random learning selection described earlier. Letters A thru E indicate particular tests used for a learning set. A "T" indicates the standard test set was used for the learning set - a reversal of the problem. The predictions present the number predicted correctly over the total number in the test set. The test set is composed of those gage pairs not in the learning set unless an alternate test set is noted under the LTC column and appears in parenthesis.

Fourteen time domain features predicted 80% correctly for the test set (No. 13) and 97% correctly for the learning set (No. 8). A prediction based on the same data used in learning set can be viewed as a self consistency check and is expected to be better than the prediction of unseen data. This self consistency check can be useful in the identification of odd data, however.

Fourteen frequency domain features predicted 94% correctly for the test set (No. 22) and 91.4% correctly for the learning set (No. 163).

The combination of 28 time and frequency features already considered predicted 85.7% correctly for the test data (No. 23) and 100% correctly for the learning data (No. 166).

These results are encouraging for two main reasons. First the prediction results are quite good and help show the merit of the pattern recognition procedures and secondly the success of the very simple frequency features indicates the value of extracting information from this description of the data.

Several other basic exercises are illustrated in Table 5. Test no. 19 shows 100% prediction results when all the data is used in the learning and test sets. This is another self consisting check. Tests no. 16 and 17 illustrate the ability to learn from one event of each class to predict the remaining events. While the overall prediction rate varied from an acceptable 69 to 78.6% the class 2 (berm) prediction for test no. 17 was a poor 21%. Such poor results however can be a clue of some anamoulous character in the problem set up. In this case event D (PHG 4) could be questionable.

Test no. 18 repeats test no. 17 except that each feature was normalized in the belief that numerically large features might dominate numerically small features in the FLD procedure. As seen in Table 5 the prediction results remained nearly unchanged.

Test no. 15 is a repeat of test no. 23 involving the 14 time and 14 frequency features except the standard test and learning sets were reversed. This

produced a 68.6 correct prediction compared to the earlier 85.7% and shows the FLD sensitivity to the data particularly for small data sets.

Single Feature Study

An evaluation of individual features was made by applying the FLD prediction procedure using only one feature at a time. While this does not make use of the FLD ability to operate in multi-dimensional feature space it did allow a ranking of features. Table 6 lists the results for the better performing features which are identified in Table 7. It is interesting that features from the time, frequency, and cepstrum domains are among the better performers. The histogram (Fig. 13) depicting the predicting results for all 60 features show that the majority, at least when used independently, are not very good. Features that are very good or very bad predictors are the most interesting. The case of a bad predictor may indicate a basic difference between the learning and test data sets. Another possibility is that the prediction process is essentially a coin flip. In this case the distribution around 50% prediction success occasionally produces very bad or good results.

Vertical Frequency Feature Study

Vertical gage features were thought to be more sensitive to the no berm/berm difference due to the physical fact that airblast loading would be present in one case and not another. Frequency domain features were selected based on their probability density functions and comparison of statistical factors such as mean separation and the ratio of the mean to the standard deviation.

Table 8 lists selected results from this study. Table 9 lists the prediction results versus feature number to show the effects of combining features. As seen in this table, adding additional features does not guarantee significantly improved results.

Good Feature Combinations

Features judged to be "good" based on the single feature results and on high correlation coefficient results were combined to illustrate the effect of combining good features. Table 10 illustrates the features and predictions involved and Table 11 lists the FLD tests.

This test was foreshortened by the success of combining features 23 and 29. Additional features could only degrade this prediction result. It is interesting

to note that feature 23 is a horizontal gage frequency power window and feature 29 is a vertical cepstrum peak. Neither of these features comes from the normal time domain view of the waveforms. Also, the vertical gages were thought to be more sensitive to the physical difference due to berming the explosive and yet a horizontal gage provides critical classification information. The cepstrum feature was developed for the spallation classification problem (Ref. 1) and its merit in this problem comes as a surprise.

Poor Feature Combinations

Combining good features to produce superior prediction results, as just demonstrated, is an easily accepted concept. A more difficult concept is that combining sets of poor and bad features can produce good results. Table 12 presents predictions based on a set of 12 frequency based features (No. 20) producing a 54.3% prediction success and a set of 2 frequency based features (No. 21) producing a 68.6% prediction success. These individual results are not noteworthy. In a two class problem, a 50% prediction success could be obtained by flipping a coin. However, combining these two sets of poor features into a 14 feature set (No. 22) produced a 94.3% prediction success.

Dissimilar Features

Individually "good" features were combined with other "good" features based on a low correlation coefficient between the features. The concept illustrated here is that adding additional dimensions to the FLD can either add new or reinforce existing feature information. For combinations based on a low correlation coefficient new information should be the type that is added.

Tables 13 and 14 illustrate this concept. Note in this case that features 56 and 58 are from the time domain.

FLD Modification Tests

Three modifications, described earlier, were made to the FLD. The first modification was a minor mathematical correction not expected to have much impact. The other modifications could have a significant impact on the operation of the FLD, particularly in cases involving small data sets and outliers.

In tests involving the same 14 time and 14 frequency features as the Basic Studies Tests FLD/MOD2 produced similar prediction results and FLD/MOD3 produced poorer prediction results. In tests involving single features both FLD/MOD2 and

FLD/MOD3 produced slightly improved prediction results. In the single feature case the feature projection portion of the FLD procedure is effectively inoperative, however.

X. CONCLUSIONS AND RECOMMENDATIONS

Pattern recognition error rates may be due to

- errors in the data and/or poor experimental strategies
- inadaquate signal processing
- suboptimal features
- suboptimal discriminant functions.

Fortunately "good" solutions can be found doing less than optimal jobs in any one area. In the extreme case doing a particularly good job in one area, such as feature selection, may make significant efforts in other areas unnecessary.

This project has demonstrated the merit of applying new tools to the analysis of civil engineering impulsive loading research results. These tools are predicted to produce both improved analysis economy and enhanced phenomena understanding due to the ability to more completely address and understand the information content of measured waveforms. Specifically this project has:

- demonstrated the ability of computer-automated signal analysis procedures to extract waveform information and feed it to a pattern recognition processor.
- shown that pattern recognition is a useful technique to classify waveforms and to help determine the relative value of features and feature combinations.
- shown that important information exists in the measured waveforms in other than the time domain. In fact the information residing in these other domains may be less subject to some of the errors that tend to affect time domain data. An example of this is the DC offset problem in the time domain has little impact in the frequency domain.
- shown initial work in the identification and combination of good features. This seems to be a fundamental area that needs additional effort. Included in this area is the ability to develop effective new features that are nonlinear combinations of other features.
- proposed modifications to the FLD that may assist when using data that contains outlier points. Modification 2 substituting the population median for the mean is a simple and appealing change.

In the course of this investigation several related issues became evident that merit mention. Some of these address the peculiar data characteristics found in this research area and addressed in the body of this report.

1. Small data bases can be partly addressed by improved experimental design procedures which consider the application of more sophisticated analysis procedures such as statistics and pattern recognition. Establishment of and easy access to archival data bases would allow the development of the quantity of data needed for some studies.

2. Standard time domain features have found favor because they are easily understood and easily imput to finite element predictive computer codes. As new useful non time domain features are discovered their physical significance will have to be understood and means developed to utilize these features in computer codes.

REFERENCES

1. Carson, J.M., "Application of Signal Analysis and Pattern Recognition to Study Blast Induced Ground Motion," AFOSR-TR-84-0581, March 1984.

2. Stump, B.W. and Reinke, R.E., "Spall-Like Waveforms Observed in High-Explosive Testing in Alluvium," AFWL-TR-82-15, September 1982.

3. Stump, B.W. and Reinke, R.E., "Spall Observations and Mechanisms in Alluvium," Journal of Geophysical Research, vol 89, no. B13 pages 11,495 - 11,506, December 10, 1984.

4. Babcock, S.M., "Pre-Hybrid Gust Phase I Quick Look Report," CERF AG-26, Air Force Weapons Laboratory, Kirtland Air Force Base, New Mexico, December 1979.

5. Schneider, J.F., "Signal Analysis in Instrumentation," AFWL-DR-TN-78-004, Air Force Weapons Laboratory, Kirtland Air Force Base, New Mexioc, May 1978.

6. Kemerait, R.C., and Childers, D.G., "Signal Detection and Extration by Cepstrum Techniques," IEEE Transactions on Information Theory, November 1972.

7. Duda, R.O. and Hart, P.E., Pattern Classification and Scene Analysis, Wiley Interscience, 1973.

8. Rose, J.L., Avioli, M.J., and Tapides, M.E., "Utilization of a Fisher Linear Discriminant Function in IGSCC Detection," CSNI Specialist meeting, Brussels, Belguim, May 1986.

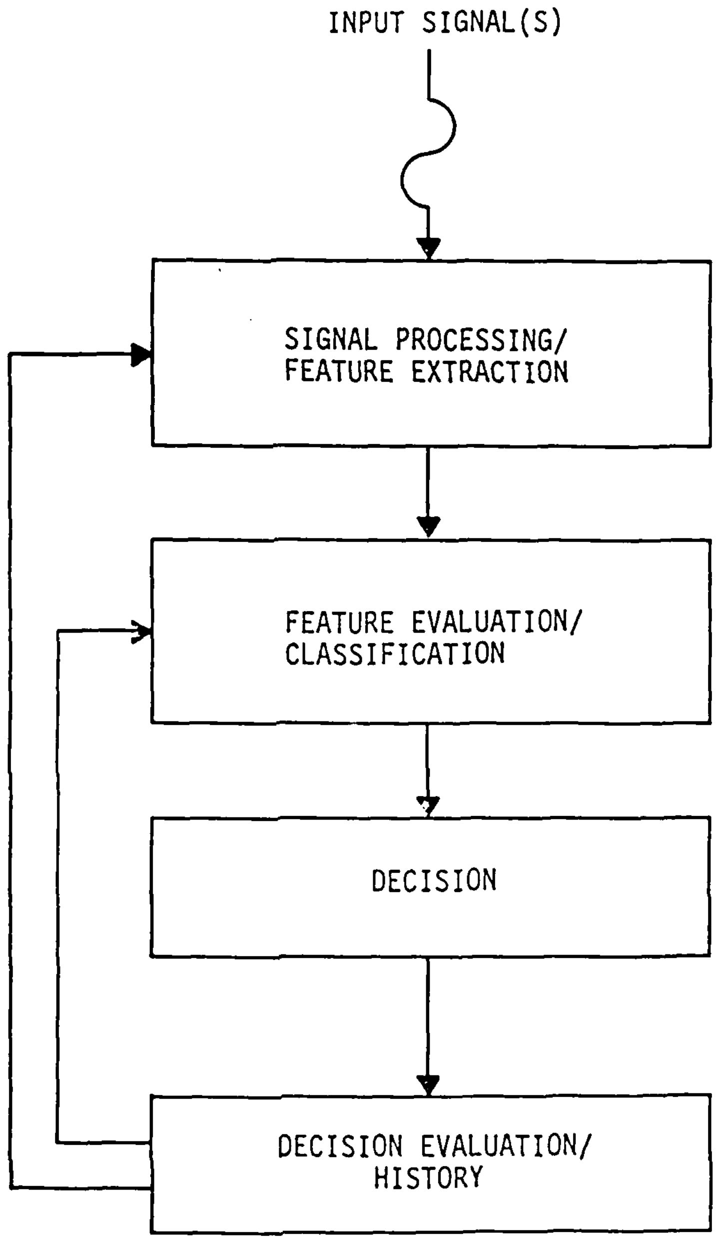


Figure 1. The information processing process.

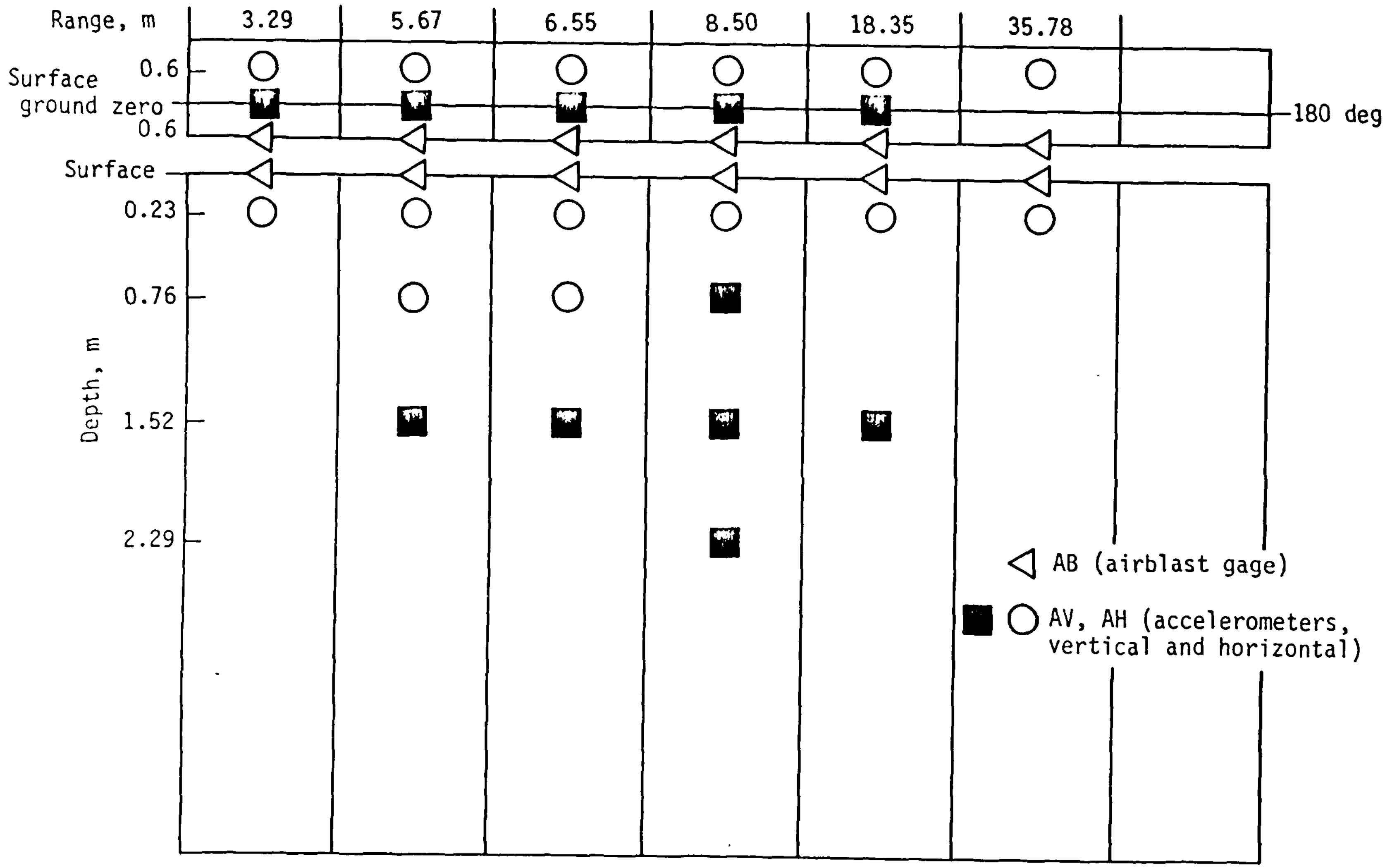
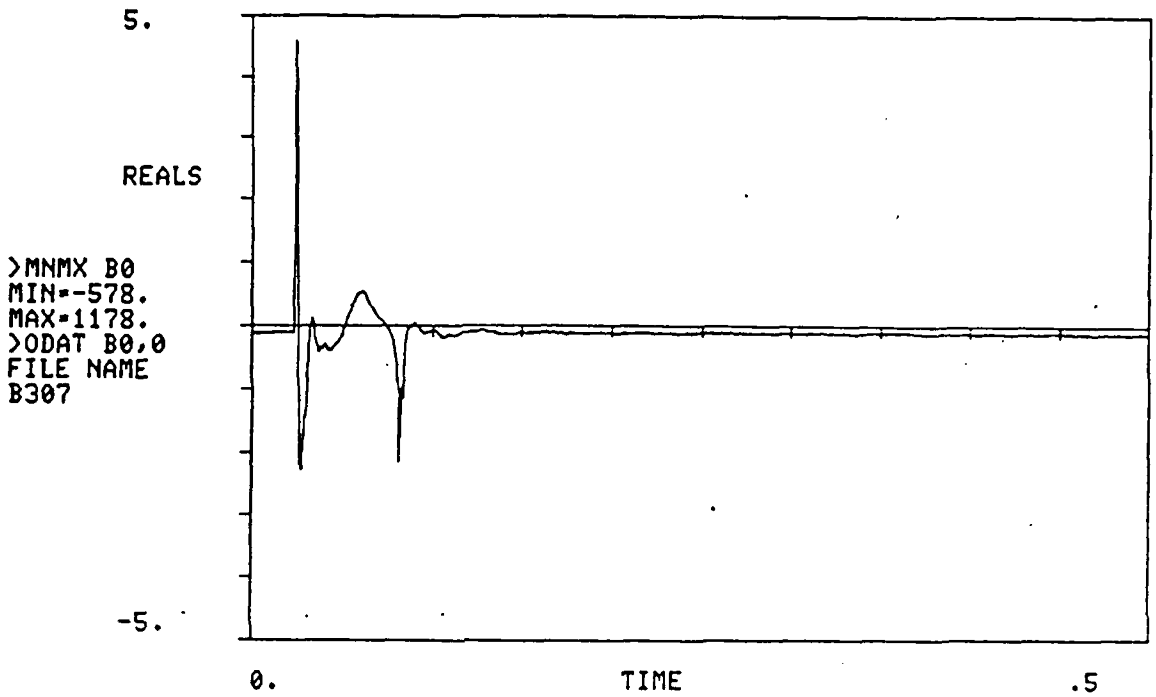
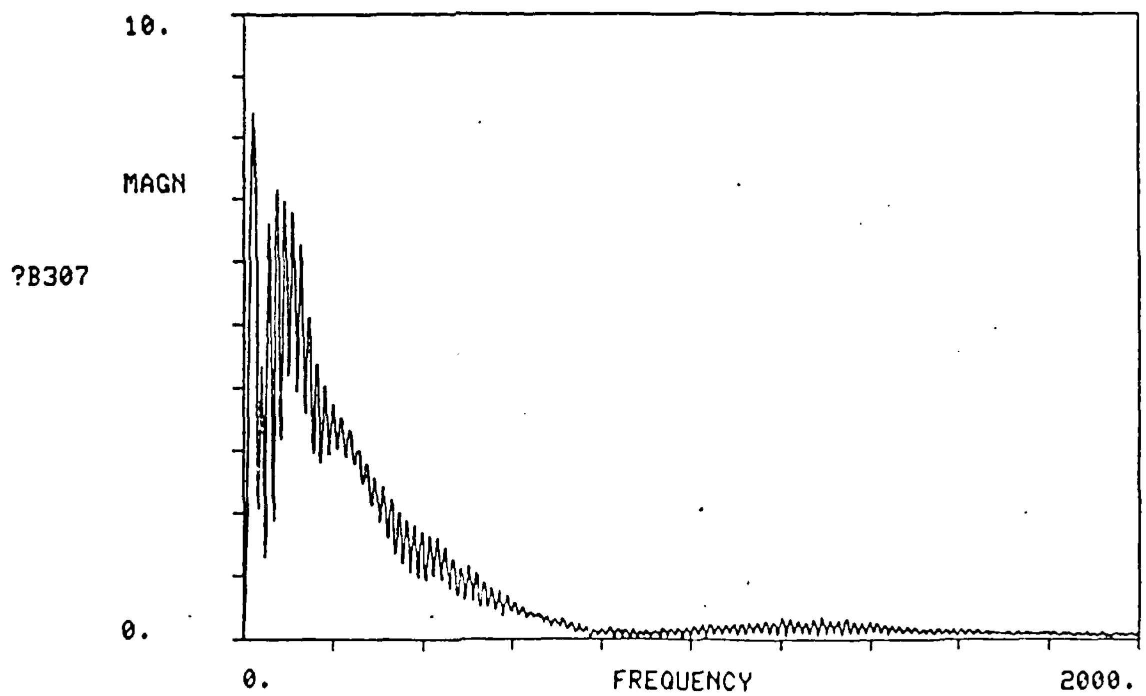


Figure 2. Gage layout for PHG I-1 through I-5.



(a) Acceleration-time record.



(b) Frequency content of Figure 3a.

Figure 3. Signal acquisition check.

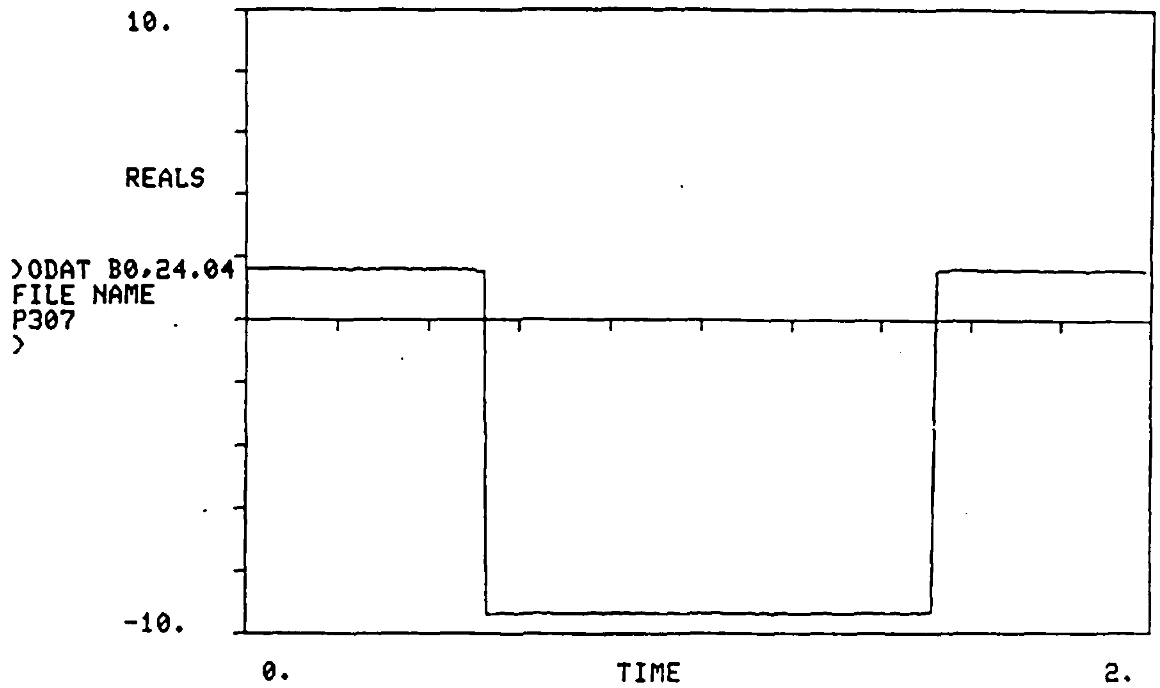


Figure 4. Transducer calibration record.

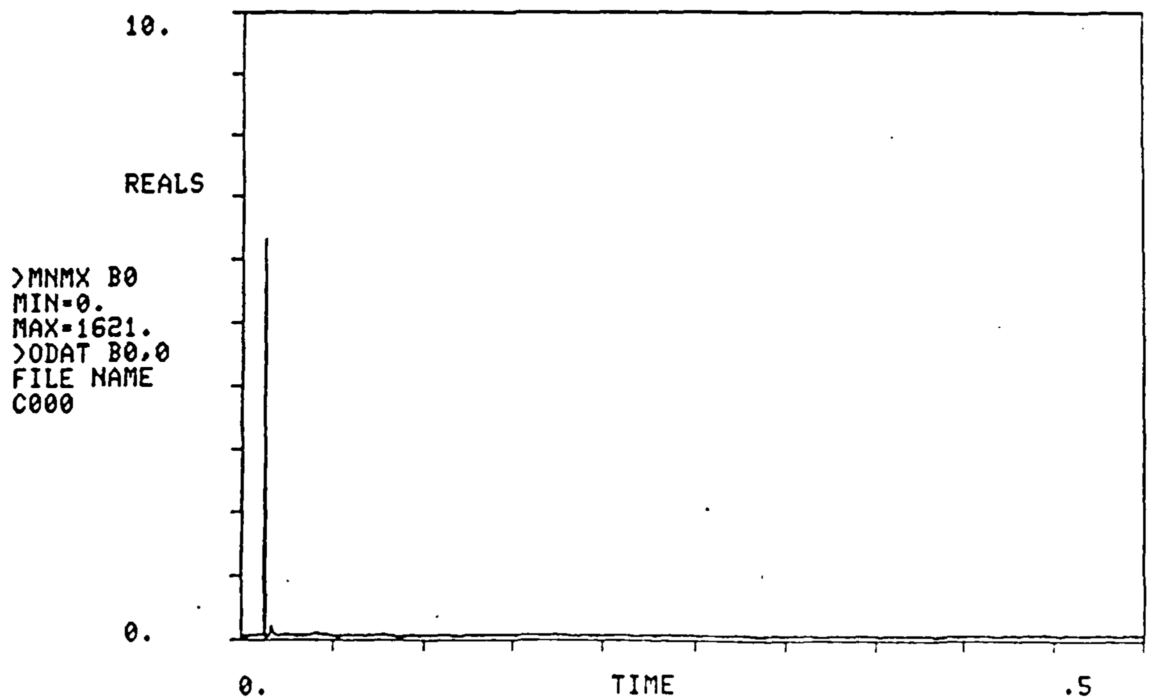


Figure 5. Explosive event fire pulse.

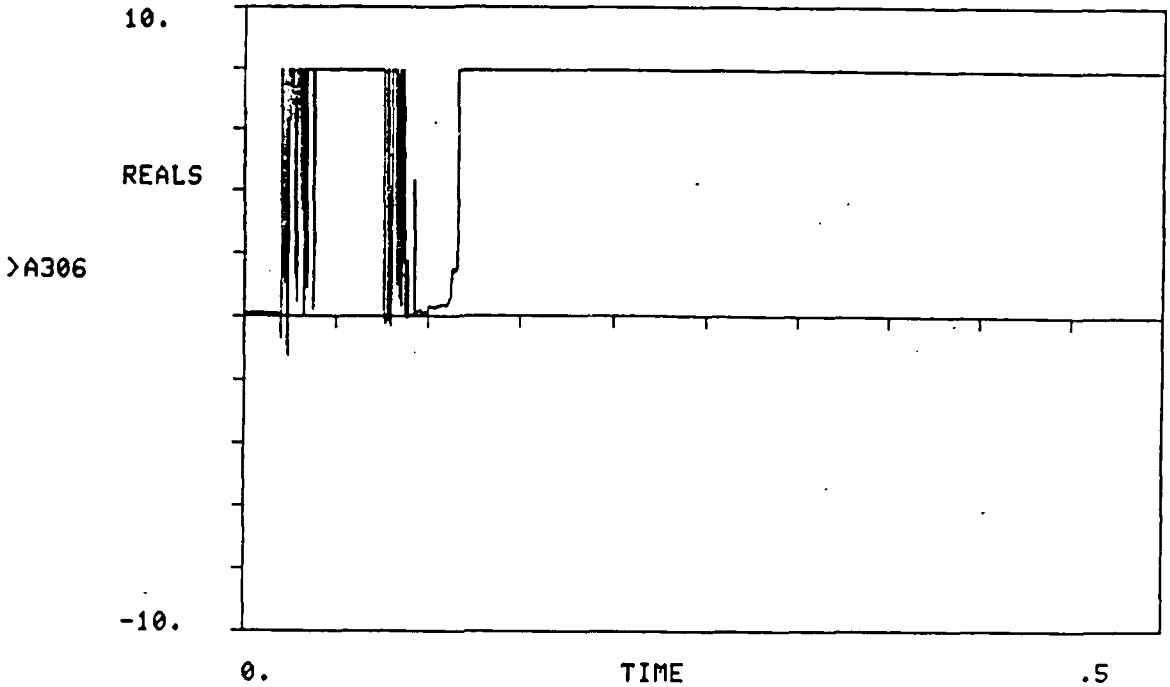


Figure 6a. A306 poor data record--probable gage failure.

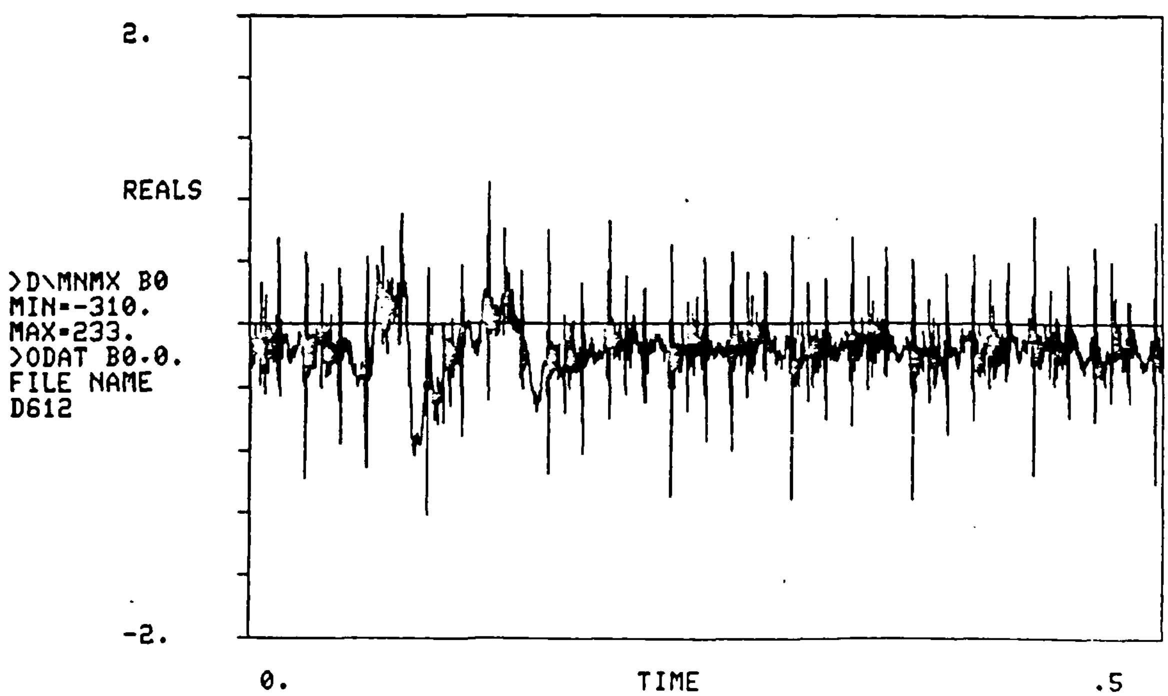


Figure 6b. D612 poor data record--low S/N ratio.

D A T A S E T

C L A S S 1   C L A S S 2

X_1 , X_2 . . . X_{n1}, X_{n1+1} . . . X_{n1+n2}

A11 (feature 1) ___ ___ ___ ___ ___ ___

A21 (feature 2) ___ ___ ___ ___ ___ ___

.

.

.

Ad1 (feature d) ___ ___ ___ ___ ___ ___

DATA SET (n samples)

- n samples are divided into subsets (classes) of n_1 and n_2 samples ($n = n_1 + n_2$)
- each sample (X_1, X_2, . . . X_n) is a vector of dimension d (feature vector)
- each sample may result from a single instrument, multiple instruments, or an entire test depending on experiment design

Figure 7. Data set structure

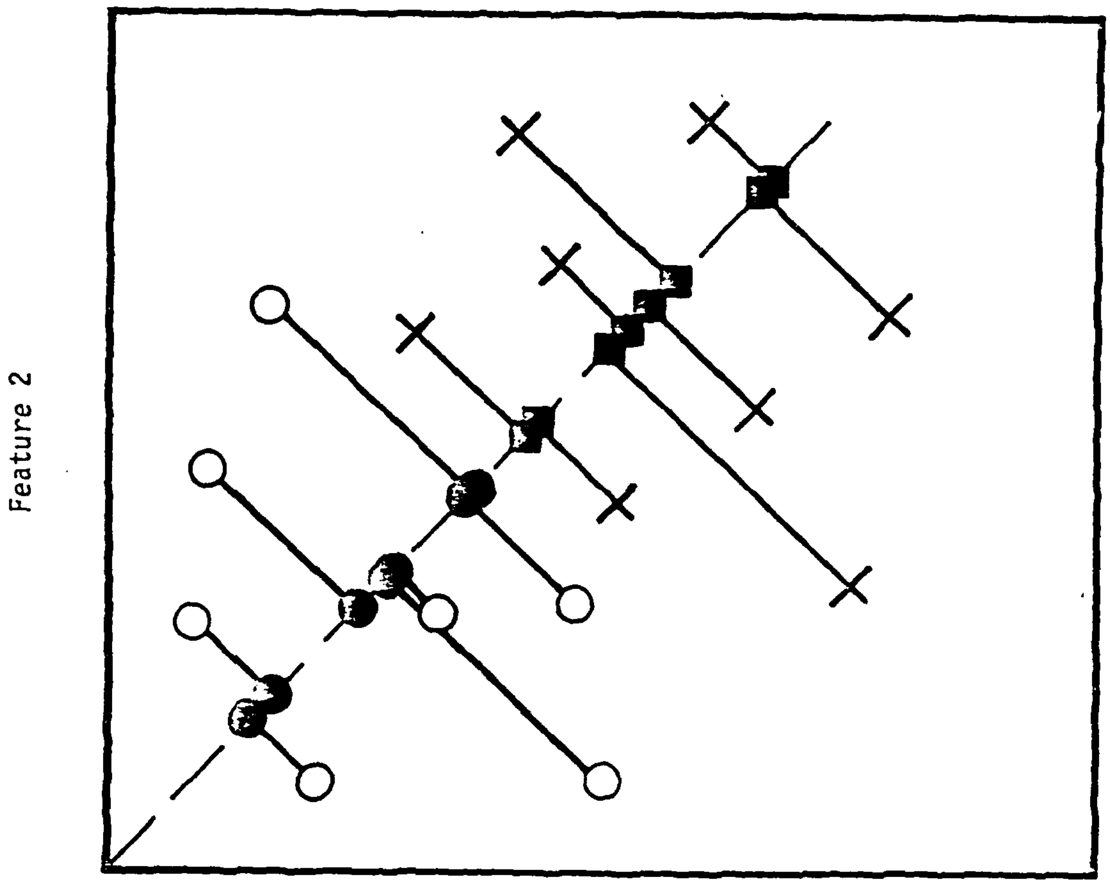


Figure 8. Illustration of Fisher's linear discriminant.

```
FEATURE #      50
# OF CLASS 1 VECTORS:      31
# OF CLASS 2 VECTORS:      14
YBAR=  0.3366E-01 S=  0.2241E-01
YBAR=  0.6433E-01 S=  0.3267E-01
```

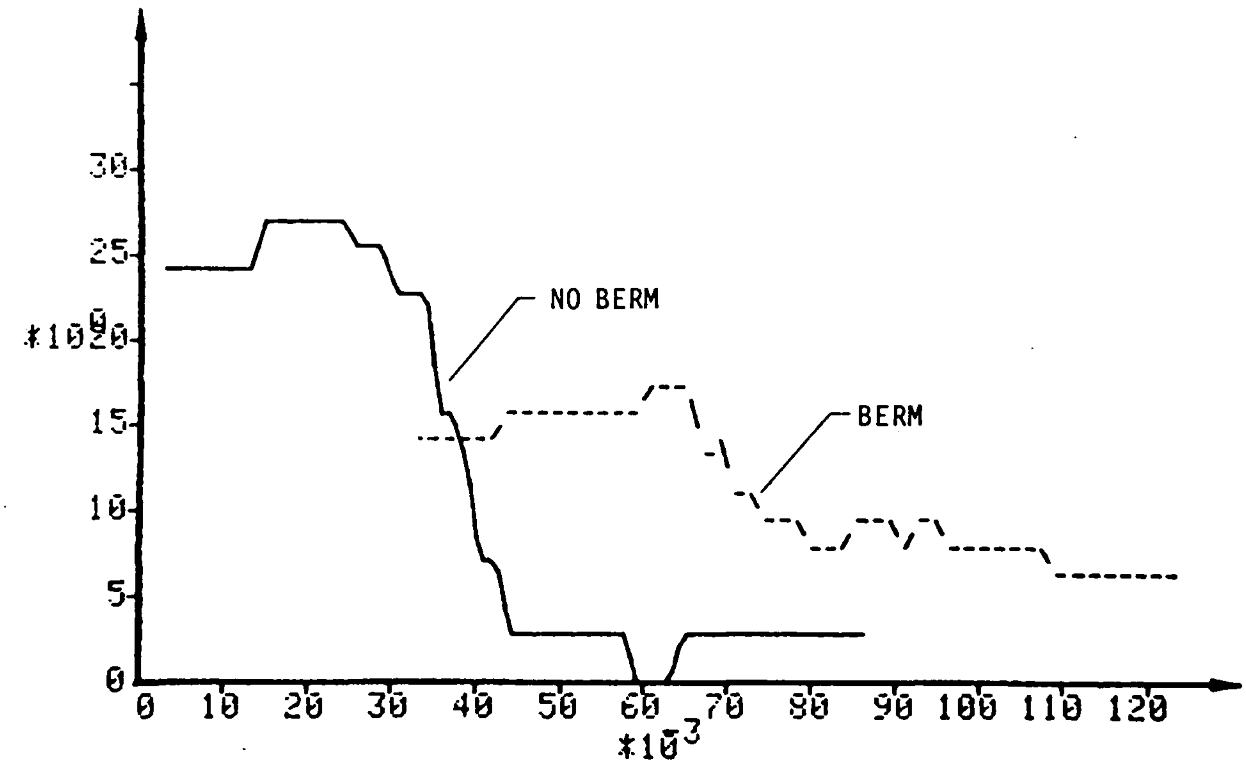


Figure 9. Example probability density function (Feature 5: vertical acceleration, time of maximum peak.

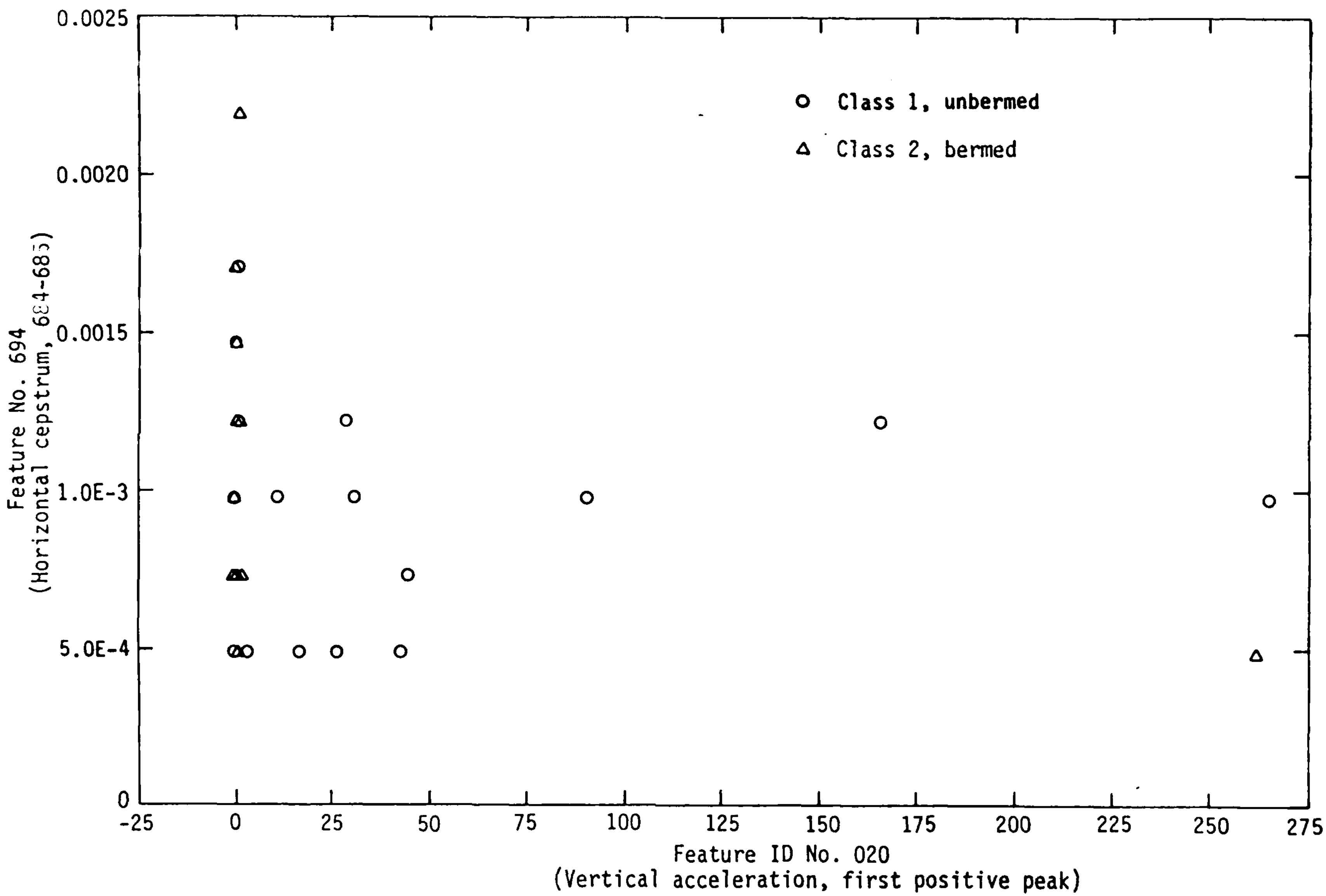


Figure 10. Example scatter plot, 020 (vertical acceleration, first positive peak), 694; horizontal cepstrum, 684-685.

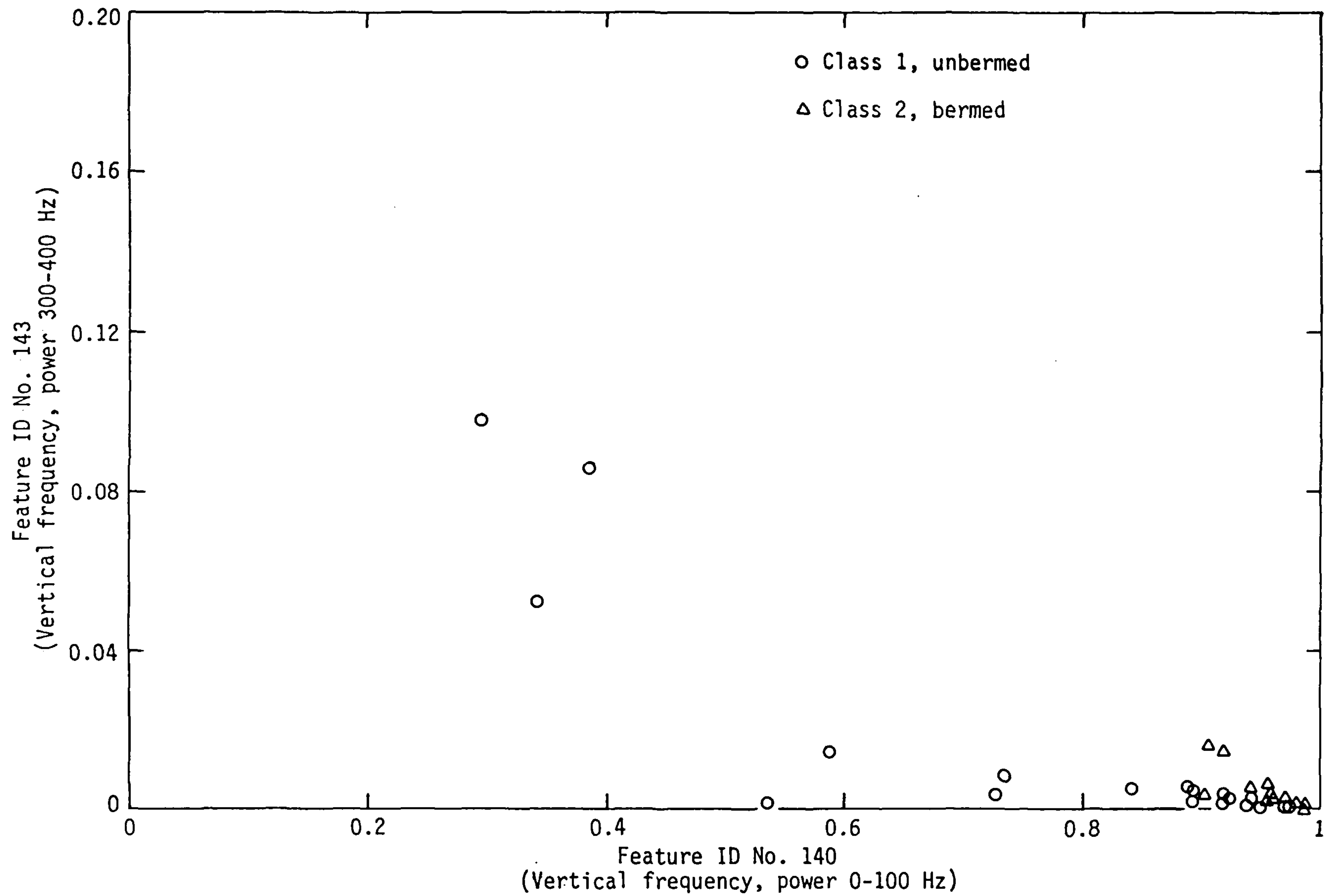


Figure 11. Example scatter plot.

	RANGE (M)						
D :		3.3	5.7	6.7	8.5	18.4	35.8
E :	.23	3 A	3 A	3 A	3 A	3 A	3 A
P :		0 B	0 B	0 B	0 B	0 B	1 B
T :		1 C	3 C	5 C	7 C	9 C	1 C
H :		/ D	/ D	/ D	/ D	/ D	/ D
(M) :		2 E	4 E	6 E	8 E	10 E	12 E
	.76		3 A	3 A	6 A		
			1 B	1 B	0 B		
			3 C	5 C	9 C		
			/ D	/ D	/ D		
			14 E	16 E	10 E		
	1.56		6 A	6 A	6 A	6 A	
			0 B	0 B	0 B	1 B	
			1 C	5 C	7 C	1 C	
			/ D	/ D	/ D	/ D	
			2 E	6 E	8 E	12 E	
	2.29			6 A			
				0 B			
				3 C			
				/ D			
				4 E			

Figure 12. Gage grid with standard learning set selections circled

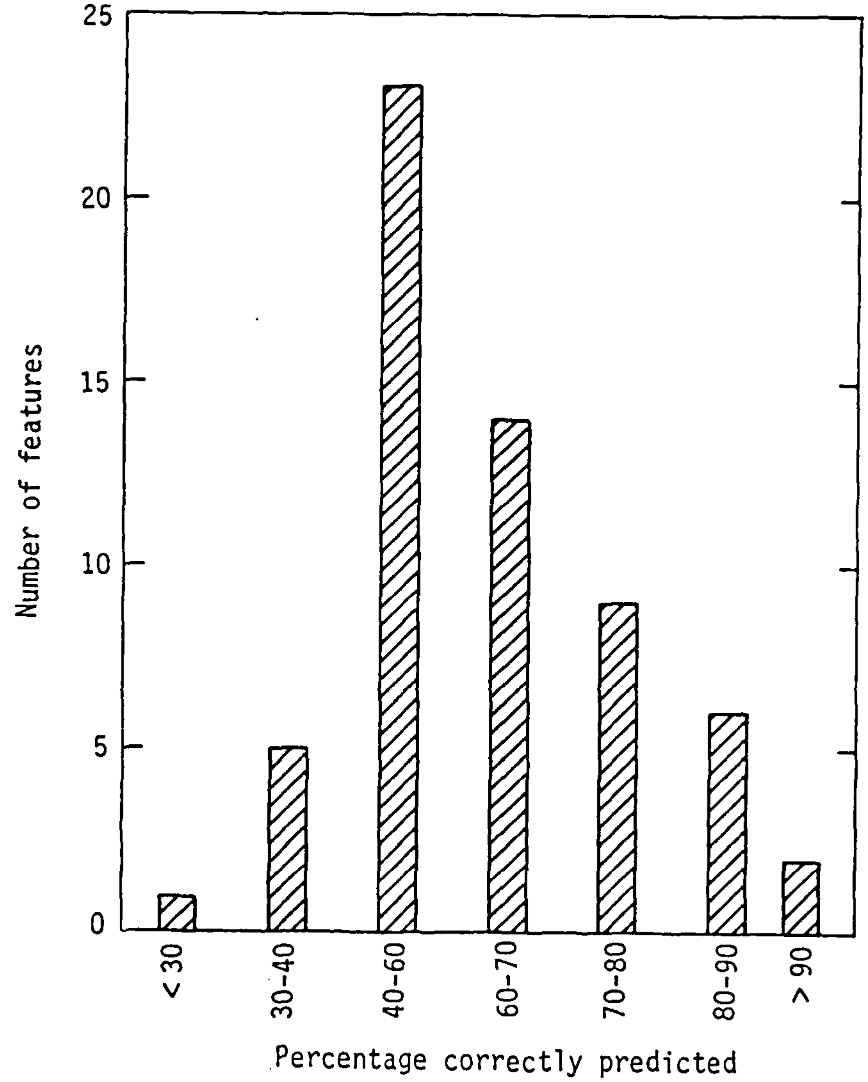


Figure 13. Single feature study, prediction success summary.

TABLE 1. PHG I SERIES OUTLINE

Event	Airblast	Bermed	Explosive weight, kg			Gage placement			Canister		Pit/berm size			Number accelerometer gages	
			13.6	39	116	Soil	Concrete sand	Sand buried cable	AL	ST	SM	M	L	H	V
1	X		X			X			–	26	X			14	14
2	X		X			X			–	26	X			14	14
3	X		X			X			–	26	X			14	14
4		X	X			X			–	26		X		14	14
5		X	X			X			–	26		X		14	14
6		X		X			X		18	5			X	20	20
7	X				X		X		18	5			X	20	20
8			X					X	27	–			X		
9	X				X			X	27	–			X	20	20

LEGEND

AL = aluminum
ST = steel
SM = small
M = medium
L = large
H = horizontal
V = vertical

Premise: Horizontal motion same; vertical differences indicates pore air; 116-kg STS TNT charge standard; all McCormick Ranch.

TABLE 2. ACCELEROMETER DATA FOR PHG I-1 THROUGH I-5

Measurement No.	Range, m	Depth, m	PHG I-1 +	PHG I-1 −	PHG I-2 +	PHG I-2 −	PHG I-3 +	PHG I-3 −	PHG I-4 +	PHG I-4 −	PHG I-5 +	PHG I-5 −
1301	3.23	0.23V	50.00	175.00	60.0	275.0	33.00	90.00	4.40	2.40	7.20Q	2.80
1302	3.23	0.23H	43.00	15.00	65.0	30.0	27.00	10.00	11.00	4.00	14.00	4.00
1303	5.67	0.23V	23.00	43.00	10.0	22.0	15.00	41.00	1.80	0.75	1.50Q	1.61
1304	5.67	0.23H	-----	-----	4.0	3.2	24.00[b]	0.25	3.10	3.00	4.70	2.80
1305	6.55	0.23V	20.00	25.00	7.5	17.0	15.00	43.00	2.00	1.00	2.20N	1.50
1306	6.55	0.23H	a	a	2.2	2.5	14.00	7.50	2.10	2.10	2.20	2.30
1307	8.50	0.23V	14.00	31.00	14.0	30.0	15.00	27.00	1.20	0.90	1.20N	1.20
1308	8.50	0.23H	5.50	5.50	10.0	6.5	8.70	3.70	0.75	1.00	1.30	1.20
1309	18.35	0.23V	5.50	11.00	5.0	9.5	7.00	9.50	0.45	0.47	0.28	0.40
1310	18.35	0.23H	2.50	1.50	2.4	1.6	1.00	1.50	0.45	0.35	0.42	0.33
1311	35.78	0.23V	3.60	3.30	3.4	3.4	2.50	2.80	0.14	0.23	0.14	0.12
1312	35.78	0.23H	0.65	1.10	0.4	1.0	0.56	0.75	0.45	0.33	0.13	0.13
1313	5.67	0.76V	2.20	2.60	----	---	7.90	7.00	2.00N	1.00	1.20N	1.20
1314	5.67	0.76H	0.87	0.67	2.2	2.7	1.70	2.50	2.50	2.30	0.76	0.26
1315	6.55	0.76V	2.70	2.60	4.4	5.2	8.00	12.00	2.00N	1.00	1.50N	1.40
1316	6.55	0.76H	0.90	1.10	1.7	1.5	1.80	1.50	1.50	1.60	2.00	1.30
1601	5.67	1.52V	3.70	2.20	4.0	2.5	4.00	2.80	2.00	1.70	2.20	2.20
1602	5.67	1.52H	1.00	2.20	1.9	2.5	1.00	2.40	1.40	1.50	1.80	2.00
1603	6.55	2.29V	3.10	1.90	3.1	1.9	3.30	1.80	1.80	1.10	2.70	1.40
1604	6.55	2.29H	1.50	1.60	2.9	1.8	2.80	1.90	1.50	1.60	1.60	2.30
1605	6.55	1.52V	3.50	2.10	3.2	2.0	3.20	2.00	1.60	1.20	1.70	1.60
1606	6.55	1.52H	1.20	2.20	1.2	1.7	1.20	1.70	0.85	0.75	0.80	1.30
1607	8.50	1.52V	2.50	2.00	2.0	2.1	1.80	1.70	1.30	0.75	1.60	0.80
1608	8.50	1.52H	1.40	2.60	1.5	2.6	1.40	2.70	0.48	0.75	0.60	1.20
1609	8.50	0.76V	3.10	3.00	3.2	3.1	3.00	3.00	1.30Q	0.80	1.30Q	0.80
1610	8.50	0.76V	1.50	1.80	1.7	2.0	1.90	2.30	0.56	0.70	0.75	0.85
1611	18.35	1.52V	1.50	0.80	1.0	0.8	0.95	0.65	0.25N	0.22	0.32	0.25
1612	18.35	1.52H	0.75	1.40	0.7	1.2	0.70	1.40	0.20N	0.15	0.20N	0.26

Note: All measurements in g unless indicated otherwise. [a]Bad data. [b]Data questionable. V = vertical gage. H = horizontal gage. Q = quantizing error. N = noisy.

(from Reference 4, Table 7a.)

TABLE 3. POOR WAVEFORMS VIA VISUAL EXAMINATION

Test/Gage	Problem
A304*	bandedge
A306	probable gage failure, TOA useful rest doubtful
D612*	very low S/N ratio
D313*	low S/N ratio

* gage normally a member of the learning set.

TABLE 4 LIST OF FEATURES

NO.	HORIZ/VERT	EXTRACTION SUBROUTINE	ID NO.	FEATURE DESCRIPTION
1	V	1	010	acceleration, time of arrival
2	V	2	020	acceleration, lst. positive peak
3	V	3	030	acceleration, lst. negative peak
4	V	4	040	acceleration, maximum peak
5	V	5	050	acceleration, time of maximum peak
6	V	6	060	velocity, lst positive peak
7	V	7	070	velocity, lst negative peak
8	H	1	510	acceleration, time of arrival
9	H	2	520	acceleration, lst positive peak
10	H	3	530	acceleration, lst negative peak
11	H	4	540	acceleration, maximum peak
12	H	5	550	acceleration, time of maximum peak
13	H	6	560	velocity, lst positive peak
14	H	7	570	velocity, lst negative peak
15	V	12	120	frequency, peak
16	V	14	140	frequency, power 0-100 HZ
17	V	14	141	frequency. power 100-200 HZ
18	V	14	142	frequency, power 200-300 HZ
19	V	14	143	frequency, power 300-400 HZ
20	V	14	144	frequency, power 400-500 HZ
21	V	14	145	frequency, power 500-600 HZ
22	H	14	620	frequency, peak
23	H	14	640	frequency, power 0-100 HZ
24	H	14	641	frequency, power 100-200 HZ
25	H	14	642	frequency, power 200-300 HZ
26	H	14	643	frequency, power 400-500 HZ
27	H	14	644	frequency, power 400-500 HZ
28	H	14	645	frequency, power 500-600 HZ
29	V	18	180	cepstrum, time of lst peak
30	V	18	181	cepstrum, time of 2nd peak

TABLE 4 LIST OF FEATURES (CONTINUED)

NO.	HORIZ/VERT	EXTRACTION SUBROUTINE	ID NO.	FEATURE DESCRIPTION
31	V	18	182	cepstrum, time of 3rd peak
32	V	18	183	cepstrum, time of 1st largest peak
33	V	18	184	cepstrum, time of 2nd largest peak
34	V	18	185	cepstrum, time of 3rd largest peak
35	V	19	190	cepstrum, 180-181
36	V	19	191	cepstrum, 181-182
37	V	19	192	cepstrum, 182-180
38	V	19	193	cepstrum, 183-184
39	V	19	194	cepstrum, 184-185
40	V	19	195	cepstrum, 185-183
41	H	18	680	cepstrum, time of 1st peak
42	H	18	681	cepstrum, time of 2nd peak
43	H	18	682	cepstrum, time of 3rd peak
44	H	18	683	cepstrum, time of 1st largest peak
45	H	18	684	cepstrum, time of 2nd largest peak
46	H	18	685	cepstrum, time of 3rd largest peak
47	H	19	690	cepstrum, 680-681
48	H	19	691	cepstrum, 681-682
49	H	19	692	cepstrum, 682-680
50	H	19	693	cepstrum, 683-684
51	H	19	694	cepstrum, 684-685
52	H	19	695	cepstrum, 685-683
53	V	8	080	velocity, peak
54	V	9	090	velocity, time of peak
55	V	10	100	displacement, peak
56	V	11	110	displacement, time of peak
57	H	8	580	velocity, peak
58	H	9	590	velocity, time of peak
59	H	10	600	displacement, peak
60	H	11	610	displacement, time of peak

TABLE 5. BASIC STUDIES - PREDICTION RESULTS

TEST NO	CLASS 1	CLASS 2	COMBINED	LEARNING / TEST	CRITERIA	FEATURES
13	18/21	10/14	28/35	L		000000000111111111122222222223333333333444444444455555555556 12345678901234567890123456789012345678012345678901234567890 ●●●●●●●●●●●●●●0000000000000000000000000000000000000000000000
22	19/21	14/14	33/35	L		000000000111111111122222222223333333333444444444455555555556 12345678901234567890123456789012345678012345678901234567890 00000000000000●●●●●●●●●●●●●●00000000000000000000000000000000
23	18/21	12/14	30/35	L		000000000111111111122222222223333333333444444444455555555556 12345678901234567890123456789012345678012345678901234567890 ●●●●●●●●●●●●●●●●●●●●●●●●●●●●00000000000000000000000000000000
8	21/21	13/14	34/35	L (L)		000000000111111111122222222223333333333444444444455555555556 12345678901234567890123456789012345678012345678901234567890 ●●●●●●●●●●●●●●0000000000000000000000000000000000000000000000
163	19/21	13/14	32/35	L (L)		000000000111111111122222222223333333333444444444455555555556 12345678901234567890123456789012345678012345678901234567890 00000000000000●●●●●●●●●●●●●●00000000000000000000000000000000
166	21/21	14/14	35/35	L (L)		000000000111111111122222222223333333333444444444455555555556 12345678901234567890123456789012345678012345678901234567890 ●●●●●●●●●●●●●●●●●●●●●●●●●●●●00000000000000000000000000000000

TABLE 5. BASIC STUDIES - PREDICTION RESULTS (Cont.)

TEST NO	CLASS 1	CLASS 2	COMBINED	LEARNING/TESTING CRITERIA	FEATURES
					00000000011111111112222222222333333333344444444445555555556 12345678901234567890123456789012345678012345678901234567890
19	42/42	28/28	70/70	ABCDE (ABCDE)	●●●●●●●●●●●●●●●●●●●●●●●●●●●●0000000000000000000000000000000
					00000000011111111112222222222333333333344444444445555555556 12345678901234567890123456789012345678012345678901234567890
15	17/21	7/14	24/35	T	●●●●●●●●●●●●●●●●●●●●●●●●●●●●0000000000000000000000000000000
					00000000011111111112222222222333333333344444444445555555556 12345678901234567890123456789012345678012345678901234567890
16	23/28	10/14	33/42	CE	●●●●●●●●●●●●●●●●●●●●●●●●●●●●0000000000000000000000000000000
					00000000011111111112222222222333333333344444444445555555556 12345678901234567890123456789012345678012345678901234567890
17	26/28	3/14	29/42	BD	●●●●●●●●●●●●●●●●●●●●●●●●●●●●0000000000000000000000000000000
					00000000011111111112222222222333333333344444444445555555556 12345678901234567890123456789012345678012345678901234567890
18	25/28	3/14	28/42	BD	●●●●●●●●●●●●●●●●●●●●●●●●●●●●0000000000000000000000000000000 (features normalized)

TABLE 6. SINGLE FEATURE STUDY - PREDICTION RESULTS

TEST NO	CLASS 1	CLASS 2	COMBINED	LEARNING/TESTING	CRITERIA	FEATURES
55	21/21	11/14	32/35	L		00000000011111111112222222222333333333444444444455555555556 12345678901234567890123456789012345678012345678901234567890 0000000000000000000000000000000●000000000000000000000000000
83	21/21	11/14	32/35	L		00000000011111111112222222222333333333444444444455555555556 12345678901234567890123456789012345678012345678901234567890 0000000000000000000000000000000000000000000000000000000000●
46	16/21	14/14	30/35	L		00000000011111111112222222222333333333444444444455555555556 12345678901234567890123456789012345678012345678901234567890 0000000000000000000000●000000000000000000000000000000000000
35	17/21	12/14	29/35	L		00000000011111111112222222222333333333444444444455555555556 12345678901234567890123456789012345678012345678901234567890 00000000000●00000000000000000000000000000000000000000000000
47	15/21	14/14	29/35	L		00000000011111111112222222222333333333444444444455555555556 12345678901234567890123456789012345678012345678901234567890 00000000000000000000000●00000000000000000000000000000000000
52	21/21	7/14	28/35	L		00000000011111111112222222222333333333444444444455555555556 12345678901234567890123456789012345678012345678901234567890 0000000000000000000000000000●000000000000000000000000000000

TABLE 6. SINGLE FEATURE STUDY - PREDICTION RESULTS (Cont.)

TEST NO	CLASS 1	CLASS 2	COMBINED	LEARNING / TEST	CRITERIA	FEATURES
						000000000111111111122222222223333333333444444444455555555556 12345678901234567890123456789012345678012345678901234567890
79	14/21	14/14	28/35	L		0000000000000000000000000000000000000000000000000000000●0000
						000000000111111111122222222223333333333444444444455555555556 12345678901234567890123456789012345678012345678901234567890
82	17/21	11/14	28/35	L		0000000000000000000000000000000000000000000000000000000000●0

TABLE 7. BEST FEATURES FROM SINGLE FEATURE STUDY

FEATURE	DESCRIPTION
12	H, acceleration, time of maximum peak
23	H, frequency, power 0-100 HZ
24	H, frequency, power 100-200 HZ
29	V, cepstrum, time of 1st peak
32	V, cepstrum, time of 1st largest peak
56	V, displacement, time of peak
59	H, displacement, peak
60	H, displacement, time of peak

(H-horizontal gage, V-vertical gage)

TABLE 8. VERTICAL FREQUENCY FEATURE STUDY - PREDICTION RESULTS.

TEST NO	CLASS 1	CLASS 2	COMBINED	LEARNING/TEST CRITERIA	FEATURES
38	18/21	6/14	24/35	L	00000000011111111112222222222333333333344444444445555555556 12345678901234567890123456789012345678012345678901234567890 00000000000000●00000000000000000000000000000000000000000000
40	10/21	14/14	24/35	L	00000000011111111112222222222333333333344444444445555555556 12345678901234567890123456789012345678012345678901234567890 0000000000000000●000000000000000000000000000000000000000000
42	4/21	13/14	17/35	L	00000000011111111112222222222333333333344444444445555555556 12345678901234567890123456789012345678012345678901234567890 000000000000000000●0000000000000000000000000000000000000000
43	4/21	13/14	17/35	L	00000000011111111112222222222333333333344444444445555555556 12345678901234567890123456789012345678012345678901234567890 0000000000000000000●000000000000000000000000000000000000000
44	4/21	9/14	13/35	L	00000000011111111112222222222333333333344444444445555555556 12345678901234567890123456789012345678012345678901234567890 00000000000000000000●00000000000000000000000000000000000000
85	13/21	14/14	27/35	L	00000000011111111112222222222333333333344444444445555555556 12345678901234567890123456789012345678012345678901234567890 00000000000000●0●000000000000000000000000000000000000000000

TABLE 8. VERTICAL FREQUENCY FEATURE STUDY - PREDICTION RESULTS (Cont.)

TEST NO	CLASS 1	CLASS 2	COMBINED	L/T LEARNING/TESTING	CRITERIA	FEATURES
						00000000011111111112222222222333333333344444444445555555556 12345678901234567890123456789012345678012345678901234567890
95	12/21	14/14	26/35	L		00000000000000●0●000●00000000000000000000000000000000000000
						00000000011111111112222222222333333333344444444445555555556 12345678901234567890123456789012345678012345678901234567890
99	13/21	14/14	27/35	L		00000000000000●0●0●0●00000000000000000000000000000000000000
						00000000011111111112222222222333333333344444444445555555556 12345678901234567890123456789012345678012345678901234567890
104	15/21	14/14	29/35	L		00000000000000●0●0●●000000000000000000000000000000000000000

TABLE 9. VERTICAL FREQUENCY FEATURE STUDY - PREDICTION SUCCESS.

Feature(s)	% Predicted Correctly
15	68.8
17	68.8
19	48.6
20	48.6
21	37.1
15, 17	77.0
15,17,21	74.3
15,17,19,21	77.0
15,17,19,20,21	82.8

TABLE 10. GOOD FEATURE COMBINATIONS - PREDICTION SUCCESS

Feature(s)	% Predicted Correctly
23	85.7
29	80.0
23,29	100.0
23,29,5,24	94.3

TABLE 11. GOOD FEATURE COMBINATIONS - PREDICTION RESULTS

TEST NO	CLASS 1	CLASS 2	COMBINED	LEARNING/TEST	CRITERIA	FEATURES
						00000000011111111112222222222333333333344444444445555555556 12345678901234567890123456789012345678012345678901234567890
46	16/21	14/14	30/35	L		
						0000000000000000000000●000000000000000000000000000000000000
						00000000011111111112222222222333333333344444444445555555556 12345678901234567890123456789012345678012345678901234567890
52	21/21	7/14	28/35	L		
						0000000000000000000000000000●000000000000000000000000000000
						00000000011111111112222222222333333333344444444445555555556 12345678901234567890123456789012345678012345678901234567890
136	21/21	14/14	35/35	L		
						0000000000000000000000●00000●000000000000000000000000000000
						00000000011111111112222222222333333333344444444445555555556 12345678901234567890123456789012345678012345678901234567890
140	20/21	13/14	33/35	L		
						0000●00000000000000000●●0000●000000000000000000000●0●000000

TABLE 12. POOR FEATURE COMBINATIONS - PREDICTION RESULTS

TEST NO	CLASS 1	CLASS 2	COMBINED	LEARNING/TESTING	CRITERIA	FEATURES
						000000000111111111122222222223333333333444444444455555555556 12345678901234567890123456789012345678012345678901234567890
20	12/21	7/14	19/35	L		000000000000000●●●●●●0●●●●●●0000000000000000000000000000000
						000000000111111111122222222223333333333444444444455555555556 12345678901234567890123456789012345678012345678901234567890
21	18/21	6/14	24/35	L		00000000000000●000000●0000000000000000000000000000000000000
						000000000111111111122222222223333333333444444444455555555556 12345678901234567890123456789012345678012345678901234567890
22	19/21	14/14	33/35	L		00000000000000●●●●●●●●●●●●●●0000000000000000000000000000000

TABLE 13. DISSIMILAR FEATURES - PREDICTION RESULTS

TEST NO	CLASS 1	CLASS 2	COMBINED	LEARNING/TEST	CRITERIA	FEATURES
81	14/21	11/14	25/35	L		00000000011111111112222222222333333333344444444445555555556 12345678901234567890123456789012345678012345678901234567890 00000000000000000000000000000000000000000000000000000000●00
79	14/21	14/14	28/35	L		00000000011111111112222222222333333333344444444445555555556 12345678901234567890123456789012345678012345678901234567890 000000000000000000000000000000000000000000000000000000●0000
143	21/21	14/14	35/35	L		00000000011111111112222222222333333333344444444445555555556 12345678901234567890123456789012345678012345678901234567890 000000000000000000000000000000000000000000000000000000●0●00

TABLE 14. DISSIMILAR FEATURES - PREDICTION SUCCESS

Feature(s)	% Predicted Correctly
56	80.0
58	71.4
56,58	100.0

END

12-86

DTIC